



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/703,174	10/31/2000	Charu C. Aggarwal	YOR920000430US1	7445
7590	04/26/2007		EXAMINER	
William E Lewis Ryan Mason & Lewis LLP 90 Forest Avenue Locust Valley, NY 11560			HILLERY, NATHAN	
			ART UNIT	PAPER NUMBER
			2176	
SHORTENED STATUTORY PERIOD OF RESPONSE	MAIL DATE	DELIVERY MODE		
3 MONTHS	04/26/2007	PAPER		

Please find below and/or attached an Office communication concerning this application or proceeding.

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

Office Action Summary	Application No.	Applicant(s)
	09/703,174	AGGARWAL ET AL.
	Examiner Nathan Hillary	Art Unit 2176

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).

Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) Responsive to communication(s) filed on 22 February 2007.

2a) This action is FINAL. 2b) This action is non-final.

3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) Claim(s) 1-27 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5) Claim(s) _____ is/are allowed.

6) Claim(s) 1-27 is/are rejected.

7) Claim(s) _____ is/are objected to.

8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) The specification is objected to by the Examiner.

10) The drawing(s) filed on _____ is/are: a) accepted or b) objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a) All b) Some * c) None of:

1. Certified copies of the priority documents have been received.
2. Certified copies of the priority documents have been received in Application No. _____.
3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) Notice of References Cited (PTO-892)
2) Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____

4) Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____
5) Notice of Informal Patent Application
6) Other: _____

DETAILED ACTION

1. This action is responsive to communications: RCE filed on 2/22/07.
2. Claims 1 – 27 are pending in the case. Claims 1, 10, and 19 are independent.

Continued Examination Under 37 CFR 1.114

3. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 2/22/07 has been entered.

Claim Rejections - 35 USC § 102

4. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

5. Claims 1 – 8, 10 – 17 and 19 – 26 are rejected under 35 U.S.C. 102(b) as being anticipated by Chakrabarti et al. (Focused Crawling: A New Approach to Topic-specific Web Resource Discovery) [as cited by applicant].

6. **Regarding independent claim 1**, Chakrabarti et al. teach that *keyword search is used to locate an initial set of pages (using a giant crawl and index)* (p 6, section 2.2,

last paragraph), which meet the limitation of **initially retrieving one or more documents from the information network that satisfy a user-defined predicate, wherein the initial document retrieval operation is performed without assuming a specific model of a linkage structure such that the initial document retrieval operation retrieves the one or more documents without assuming that a relationship exists between a feature of a first one of the one or more documents and a feature of at least another one of the one or more documents that links to the first one.**

7. Chakrabarti et al. teach that while fetching a document, the above formulation is used to find the leaf node with the highest probability. If some ancestor has been marked good we allow future visitation of URLs found on the document, otherwise the crawl is pruned there (p 9, section Hard focus rule), which meet the limitation of **collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed and using the collected statistical information to automatically determine further document retrieval operations to be performed in accordance with the information network**, since the probabilities are calculated to find the “best” leaf node, the ancestors are analyzed to determine if they are good, and then based on that finding future visitations are allowed (p 9, section Hard focus rule). It should be noted that the *probabilities* of Chakrabarti et al. are equivalent to the claimed **statistical information**.

8. Chakrabarti et al. teach that a focused crawler is an example-driven automatic porthole-generator. We feel that the ability to focus on a topical subgraph of the Web, as

in this paper, together with the ability to browse communities within that subgraph, will lead to significantly improved Web resource discovery (p 3, last paragraph before Section 2), which meet the limitation of **wherein the statistical information-using step further comprises learning a linkage structure from at least a portion of the collected statistical information with each successive document retrieval operation such that the learned linkage structure is available for use in performing subsequent document retrieval operations requested by a user.**

9. It should be noted that the *porthole*, which is a *subgraph of the Web*, generated by the *focused crawler* of Chakrabarti et al. is equivalent to the claimed **linkage structure** that is learned. It should further be noted that the generation of a porthole or specialized link structure (p 20, last paragraph) is equivalent to the claimed **learning a linkage structure**.

10. **Regarding dependent claim 2**, Chakrabarti et al. teach that Query construction is not a one-time investment, because as pages on the topic are discovered, their additional vocabulary must be folded in manually into the query for continued discovery (p 7, lines 4 – 6), which meet the limitation of **the user-defined predicate specifies content associated with a document**. It should be noted that the *additional vocabulary of pages on the topic* of Chakrabati et al. is equivalent to the claimed **content associated with a document**.

11. **Regarding dependent claims 3 and 4**, Chakrabarti et al. teach that pages that are examples associated with a topic can be preprocessed as desired by the system. The user's interest is characterized by a subset of topics that is marked good. No good topic is an ancestor of another good topic. Ancestors of good topics are called path topics. Given a Web page, a measure of its relevance must be specified to the system (p 8, lines 9 – 14), which meet the limitation of **the statistical information collection step uses content of the one or more retrieved documents** and that **the statistical information collection step considers whether the user-defined predicate has been satisfied by the one or more retrieved documents**, since a determination is made about the ancestors and preprocessed pages are used, which are equivalent to the claimed **one or more retrieved documents**. It should be noted that the *topic* of Chakrabarti et al. is equivalent to the claimed **content and predicate**.

12. **Regarding dependent claims 5 and 6**, Chakrabarti et al. teach that we have presented evidence in this section that focused crawling is capable of steadily collecting relevant resources and identifying popular, high-content sites from the crawl, as well as regions of high relevance, to guide itself. It is robust to different starting conditions, and finds good resources that are quite far from its starting point. In comparison, standard crawlers get quickly lost in the noise, even when starting from the same URLs (p 20, Section 4.8 and p 18, Figure 9), which meet the limitation of **the collected statistical information is used to direct further document retrieval operations toward documents which are similar to the one or more retrieved documents that also**

satisfy the predicate, and that the collected statistical information is used to direct further document retrieval operations toward documents which are more likely to satisfy the predicate than would otherwise occur with respect to document retrieval operations that are not directed using the collected statistical information, since the focused crawling of Chakrabarti et al. utilizes statistical information (p 3) and compares their crawler to other crawlers and outlines the other's shortcomings (Fig 9).

13. **Regarding dependent claim 7**, Chakrabarti et al. teach that multiple citations from a single document are likely to cite semantically related documents as well. This is why the distiller is used to identify pages with large numbers of links to relevant pages (p 8, last paragraph), which meet the limitation of **the collected statistical information is used to direct further document retrieval operations toward documents which are linked to by other documents which also satisfy the predicate**. It should be noted that the semantically related documents of Chakrabarti et al. is equivalent to the claimed **documents which are linked to by other documents which also satisfy the predicate**

14. **Regarding dependent claim 8**, Chakrabarti et al. teach that we describe a Focused Crawler, which seeks, acquires, indexes, and maintains pages on a specific set of topics that represent a relatively narrow segment of the Web. Thus, Web content can be managed by a distributed team of focused crawlers, each specializing in one or

a few topics (p 2, fourth paragraph), which meet the limitation of **the information network is the World Wide Web and a document is a web page.**

15. **Regarding claims 10 – 17 and 19 – 26**, the claims incorporate substantially similar subject matter as claims 1 – 8, and are rejected along the same rationale.

Claim Rejections - 35 USC § 103

16. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

17. Claims 9, 18 and 27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chakrabarti et al. as applied to claims 1 – 8, 10 – 17 and 19 – 26 above, and further in view of Chakrabarti et al. (Distributed Hypertext Resource Discovery Through Examples) [as cited by applicant] later referenced as Ch2 et al.

18. **Regarding dependent claim 9**, Chakrabati et al. do not explicitly teach that **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages.**

19. Ch2 et al. teach that other strategies are also known, such as, if the URL is of the form `http://host /path`, then the crawler may truncate components of path and try to fetch these URL's. If links could be traversed backward, e.g. using metadata at the server, the crawler may also fetch pages that point to the page being 'expanded.' (p 382,

Art Unit: 2176

Column 1, lines 29 – 37), which meet the limitation of **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages.**

20. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the teachings of Chakrabarti et al. with that of Ch2 et al. because such a combination would provide the users of Chakrabarti et al. with teachings of *the architecture of a hypertext resource discovery system using a relational database* (p 375, Column 1, lines 1 & 2).

18 and
21. **Regarding claims ~~18~~ 27, the claims incorporate substantially similar subject matter as claims ~~9~~ 9, and are rejected along the same rationale.**

Response to Arguments

22. Applicant's arguments filed 2/22/07 have been fully considered but they are not persuasive.

23. Applicant argues that Chakrabarti et al. do not teach **initially retrieving one or more documents from the information network that satisfy a user-defined predicate, wherein the initial document retrieval operation is performed without assuming a specific model of a linkage structure such that the initial document retrieval operation retrieves the one or more documents without assuming that a relationship exists between a feature of a first one of the one or more documents and a feature of at least another one of the one or more documents that links to**

the first one, because Chakrabarti assumes that there is a certain relationship between the content of a web page and the candidates that it links to. This is evident from the fact that Chakrabarti initiates crawling with a so-called "linkage sociology." (p 11, second paragraph).

The Office disagrees.

First, the Office finds no teaching that Chakrabarti initiates crawling with "linkage sociology" as asserted by applicant. Chakrabarti et al. teach several compelling ^{DI} examples of how their goal might be met including discovering linkage sociology on page 2 in the last paragraph.

Second, Chakrabarti et al. teach that *keyword search is used to locate an initial set of pages (using a giant crawl and index)* (p 6, section 2.2, last paragraph). There is no assumption of a relationship existing between features of one or more documents.

24. Applicant further argues that Chakrabarti et al. do not teach **initially retrieving one or more documents from the information network that satisfy a user-defined predicate, wherein the initial document retrieval operation is performed without assuming a specific model of a linkage structure such that the initial document retrieval operation retrieves the one or more documents without assuming that a relationship exists between a feature of a first one of the one or more documents and a feature of at least another one of the one or more documents that links to the first one**, because Chakrabarti discloses a method for focused crawling which includes making a decision to visit an unvisited page from the crawl frontier,

Art Unit: 2176

corresponding to an initial link structure on one or more visited pages (p 11, second and third paragraphs).

The Office disagrees.

First, Applicant points to the second paragraph of page 8 as evidence that Chakrabarti fails to teach the limitation; in contradistinction, the second paragraph of page 8 proves that Chakrabarti does in fact teach the limitation.

Specifically, Chakrabarti et al. teach that we can summarize the role of the focused crawler in the following terms. We are given a directed hypertext graph G whose nodes are physically distributed. In this paper, G is the web (p 8, second paragraph). Thus, Chakrabarti starts of with G as input, which is the web or an information network.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Nathan Hillary whose telephone number is (571) 272-4091. The examiner can normally be reached on M - F, 10:30 a.m. - 7:00 p.m.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather R. Herndon can be reached on (571) 272-4136. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Art Unit: 2176

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

NH



Doug Hutton
Primary Examiner
Technology Center 2100